



3. REDDIT

- **Release/Launch Date:** June 2005
- **Number of Active Users:** 430 million¹⁴⁰
- **Short Overview of Content Moderation Process:** Reddit has a centralised team of moderators who make up approximately 10% of Reddit's workforce. The majority of the platform's content moderation is decentralised to users who volunteer to moderate content on a particular subreddit.
- **Signatory to the EU'S Code of Conduct on Illegal Hate Speech Online?** No

¹⁴⁰ David Curry, "Reddit Revenue and Usage Statistics (2023)," Business of Apps, January 9, 2023, <https://www.businessofapps.com/data/reddit-statistics/>.

Key Developments

Reddit's initial User Agreement did not include a hate speech provision. In 2015, Reddit introduced a prohibition on content that "encourages or incites violence" or "threatens, harasses or bullies or encourages others to do so," which CEO Steve Huffman later argued amounted to an "implicit" prohibition on hate speech.¹⁴¹ Nevertheless, the policy noted that Reddit "generally provides a lot of leeway in what content is acceptable," even if the nature of this content may be "offensive."

This policy was the status quo until June 29, 2020, when the company added an explicit prohibition on hate speech (see Figure 8). This revision was likely spurred by widespread international protests against the police killing of George Floyd, as well as a June 2020 open letter addressed to Huffman and Reddit's Board of Directors, signed by over 600 of the platform's groups (representing thousands of moderators and millions of Reddit subscribers.)¹⁴² The open letter implored Huffman to "stand[] up to racism and hate . . . with real action" by "[e]nact[ing] a sitewide policy against racism, slurs, and hate speech [sic] targeted at protected groups." After the letter, Reddit revised its content policies and banned communities and users that promote hate based on identity or vulnerability.¹⁴³

Figure 8¹⁴⁴

Rule 1: Remember the human. Reddit is a place for creating community and belonging, not for attacking marginalized or vulnerable groups of people. Everyone has a right to use Reddit free of harassment, bullying, and threats of violence. Communities and people that incite violence or that promote hate based on identity or vulnerability will be banned.

The new policy banned the promotion of hate, which was not defined, based on identity or vulnerability, which was defined as: "Groups based on their actual and perceived race, color, religion, national origin, ethnicity, immigration status, gender, gender identity, sexual orientation, pregnancy, or disability. These include victims of a major violent event and their families." It is worth noting that this 2020 policy initially excluded groups of people who are in the majority from protection. The rule on hate "does not protect all groups or all forms of identity," the policy read. "For example, the rule does not protect groups of people who are in the majority or who promote such attacks of hate."¹⁴⁵ The public, including Reddit users, however, noted that certain groups targeted by hate speech may be part of the majority, such as women. Reddit quickly tackled this

¹⁴¹ Kevin Roose, "Reddit's C.E.O. on Why He Banned 'The_Donald' Subreddit," New York Times, June 30, 2020, <https://www.nytimes.com/2020/06/30/us/politics/reddit-bans-steve-huffman.html>.

¹⁴² https://www.reddit.com/r/AgainstHateSubreddits/comments/gyyqem/open_letter_to_steve_huffman_and_the_board_of/
¹⁴³ https://www.reddit.com/r/announcements/comments/hi3oht/update_to_our_content_policy/

¹⁴⁴ <https://web.archive.org/web/20200630002550/https://www.reddithelp.com/en/categories/rules-reporting/account-and-community-restrictions/promoting-hate-based-identity-or>

¹⁴⁵ <https://web.archive.org/web/20200630002550/https://www.reddithelp.com/en/categories/rules-reporting/account-and-community-restrictions/promoting-hate-based-identity-or>

problem, and by July 1, 2020, the company had removed references to the “majority”.¹⁴⁶ The revised policy explained that the rule on hate “does not protect those (groups) who promoted attacks of hate or who try to hide their hate in bad faith claims of discrimination.”¹⁴⁷ Reddit’s content policy has not changed since.

Analysis of Policy Scope

Reddit has historically been vocal in its support for uninhibited free speech and did not have an explicit hate speech prohibition for many years. In an “Ask Me Anything,” CEO and co-Founder Steve Huffman famously replied to a user’s question about slurs by saying:

“The way in which we think about speech is to separate behavior from beliefs... racism itself isn’t against the rules . . . [but] I believe the best defense . . . instead of trying to control what people can and cannot say through rules . . . is to repudiate those views in a free conversation . . . We cannot control people’s beliefs, but we can police their behaviors. And as it happens, communities dedicated to racist beliefs end up banned for violating rules we do have around harassment, bullying and violence”.¹⁴⁸

However, as mentioned above Reddit added a prohibition on the promotion of hatred in 2020, as also shown in Table 5. Though the scope of content covered aligns with Article 20 (2), except for the prohibition on statements of inferiority, Table 6 reveals that the spectrum of protected characteristics is much broader than the list referenced in the ICCPR, as it includes gender, color, immigration status, gender identity, disease, sexual orientation, pregnancy, and victims of a major event.

¹⁴⁶ Adriana Stephan, “Comparing Platform Hate Speech Policies: Reddit’s Inevitable Evolution,” Freeman Spogli Institute, Stanford Internet Observatory, <https://fsi.stanford.edu/news/reddit-hate-speech>

¹⁴⁷ <https://web.archive.org/web/20200702005456/https://www.reddithelp.com/en/categories/rules-reporting/account-and-community-restrictions/promoting-hate-based-identity-or>

¹⁴⁸ “Reddit’s 2017 transparency report and suspect account findings,” *r/announcements*, https://www.reddit.com/r/announcements/comments/8bb85p/reddits_2017_transparency_report_and_suspect/dx5go62/.

Table 5

<i>Content Explicitly Covered by Reddit's Hate Speech Policies</i>		2020	2021	2022	2023
Hate(ful) speech/ content					
Promotion of Hatred		X	X	X	X
Support for Organized Hate (Including Symbols)					
<i>On the basis of protected characteristics</i>	Incitement to or Threats of Violence	X	X	X	X
	Attacks	X	X	X	X
	Statements of inferiority or content that demeans	X	X	X	X
	Dehumanization				
	Expressions of contempt or disgust				
	Calls for exclusion or segregation	X	X	X	X
	Discrimination				
	Denying or mocking historical atrocities, or valorizing the perpetrators				
	Slurs				
	Harmful Stereotypes				
	Conspiracy Theories				
Cursing					

Table 6

Characteristics Protected in Reddit's Hate Speech Policies				
	2020	2021	2022	2023
Total	12	12	12	12
Race	X	X	X	X
Ethnicity	X	X	X	X
National Origin	X	X	X	X
Religion	X	X	X	X
Gender	X	X	X	X
Color	X	X	X	X
Immigration Status	X	X	X	X
Sex				
Gender Identity	X	X	X	X
Sexual Orientation	X	X	X	X
Age				
Disability	X	X	X	X
Disease/ Medical Condition				
Veteran Status				
Occupation				
Weight				
Pregnancy	X	X	X	X
Caste				
Victims of a Major Event	X	X	X	X
Socio-economic Status				
Culture				
Tribe				

Notes: An X indicates the company's hate speech policies covered that protected characteristics for at least one month during the given year.

Changes in Enforcement Volume

Reddit's 2020 content policy overhaul resulted in a huge purge of both content and groups, including its biggest Trump supporter community - "The Donald." This group was home to more

than 790,000 Reddit users. At the time, Reddit also banned approximately 2,000 other communities, including the leftist “Chapo Trap House” with 160,000 users.¹⁴⁹ Further, Reddit banned the “Gender Critical”, prominent among radical feminists whose views on sex and gender were deemed hateful to members of the trans-community.¹⁵⁰ In fact, Reddit’s 2020 transparency report noted the company removed 7,048 posts and 6,915 entire subreddits for hateful content.¹⁵¹

Interestingly, though the policy scope did not change again, the number of pieces of content removed for hate violations increased over the next two years while the number of subreddits banned for hate fell. According to Reddit’s own transparency reports, the company removed 39,056 posts and comments involving hateful content in 2021¹⁵² and 79,316 in 2022.¹⁵³ These increases mirrored a broader trend in increased content removals across policy categories, which Reddit attributed to evolving policies and enhanced enforcement abilities.¹⁵⁴ The number of entire subreddits banned for hateful content fell 93% in 2021,¹⁵⁵ to 467, and though it jumped back up slightly to 749 in 2022,¹⁵⁶ the number remained far lower than the 2020 level. It’s possible that the 2020 purge got rid of most hateful subreddits and new ones did not emerge to replace them, while violations of the hateful content policy continued to appear across the platform and Reddit got better at detecting and removing them. Nevertheless, the data in Reddit’s 2020, 2021, and 2022 transparency reports underscores the patterns revealed by the Facebook and Instagram enforcement data: changes in the amount of content actioned by a platform are not always closely correlated with announced changes in policy scope. Again, to truly understand the effect of scope creep in hate speech policies, better access to platform data will be necessary.

¹⁴⁹ Mike Isaac, “Reddit, Acting Against Hate Speech, Bans ‘The_Donald’ Subreddit,” *New York Times*, June 29, 2020, <https://www.nytimes.com/2020/06/29/technology/reddit-hate-speech.html>.

¹⁵⁰ David Artavia, “Reddit Cracks Down on Hate Speech by Deleting TERF, Pro-Trump Forums,” *Advocate*, June 29, 2020, <https://www.advocate.com/news/2020/6/29/reddit-cracks-down-hate-speech-deleting-terf-pro-trump-forums>.

¹⁵¹ “Transparency Report 2020,” *Reddit Inc.*, <https://www.redditinc.com/policies/transparency-report-2020>.

¹⁵² “Transparency Report 2021,” *Reddit Inc.*, <https://www.redditinc.com/policies/transparency-report-2021>.

¹⁵³ “Transparency Report 2022,” *Reddit Inc.*, <https://www.redditinc.com/policies/2022-transparency-report>.

¹⁵⁴ “Transparency Report 2022,” *Reddit Inc.*

¹⁵⁵ “Transparency Report 2021,” *Reddit Inc.*

¹⁵⁶ “Transparency Report 2022,” *Reddit Inc.*